Rényi-infinity Uniform Sampling via Algorithmic Diffusion

Yunbum Kook
Georgia Tech CS

Oct 4, 2024
University of Toronto, TCS seminar

Outline

- **Part I**: Uniform sampling over a convex body through diffusions

  - arXiv:2405.01425

- **Part II**: Warm-start generation without "TV collapse"

  - arXiv:2407.12967

Part I - Collaborators

In-and-Out: Algorithmic Diffusions for Sampling Convex Bodies

NeurIPS'24

Santosh Vempala
Georgia Tech

Matthew Zhang
University of Toronto

Uniform sampling is (maybe) all you need

Sampling from a log-concave
$d\pi \propto \exp(-V)\, dx$

Reduction →

Uniform sampling from
a convex body K

- Main subroutine in volume computation
- System biology
- ...

Uniform sampling in formulation

Problem. Let K: convex body in $\mathbb{R}^d$ and $\pi = \mathsf{Unif}(K)$. How many membership oracle queries are needed to generate a sample X whose law is ε-close to π in some D?

$$D(\mathsf{law}(X), \pi) \le \varepsilon \text{ for some probability divergence/distance } D = \mathsf{TV}, \mathsf{KL}, \chi^2 \text{ etc.}$$

Geometric random walk

Ball walk(δ)

1. Pick $z \in B_\delta(x)$
2. Move to z if $z \in K$. Stay at x o.w.

Hit-and-Run

1. Pick a uniform random line ℓ through the current point x
2. Move to a uniform random point on the chord $\ell \cap K$

Another line of log-concave sampling research

Problem (Well-conditioned log-concave sampling).

Let $\pi \propto \exp(-V)$ be a smooth unconstrained distribution with $\alpha I \preceq \nabla^2 V \preceq \beta I$ (strong convexity and smoothness of a potential V) over $\mathbb{R}^d$.

How many access to the <u>first-order oracle</u> of V are needed to generate a sample X whose law is ε-close to π?

Well-conditioned log-concave sampling

General approach for getting an implementable algorithms

1. Understand the Langevin dynamics (SDE) with stationary $\pi \propto \exp(-V)$:

$$\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t$$

2. Discretize it in time:

- Euler-Maruyama discretization
- Randomized midpoint method
- So on...

Well-conditioned log-concave sampling

Analysis

1. Establish the mixing of the Langevin dynamics in W_2, KL, χ^2 (or generally Rényi)
2. Discretization-analysis somehow preserves the mixing metric
   - Girsanov's theorem [Dalalyan and Tsybakov'12]
   - Interpolation method [Vempala and Wibisono'19]
   - Hypercontractivity [Chewi et al.'21]
   - Shifted composition rule [Altschuler and Chewi'24]

Hierarchy of probability distance/divergence

$$\mathscr{R}_\infty(\mu\|\pi) = \operatorname{esssup}\log\frac{\mathrm{d}\mu}{\mathrm{d}\pi}$$

$\downarrow$

$$\mathscr{R}_q(\mu\|\pi) := \frac{1}{q-1}\log\mathbb{E}_\mu\Big[\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\pi}\Big)^{q-1}\Big]$$ (q-Rényi divergence)

$\downarrow$

$$\mathscr{R}_2(\mu\|\pi) = \log\big(\chi^2(\mu\|\pi)+1\big)$$

$\downarrow$

$$\lim_{q\to1}\mathscr{R}_q = \mathrm{KL}$$

Pinsker $\rightarrow$

$$\mathrm{TV}(\mu,\pi) = \sup_S|\mu(S)-\pi(S)|$$

Talagrand $\rightarrow$

$$W^2(\mu,\pi) = \inf_{\Gamma(\mu,\pi)}\mathbb{E}_{(X,Y)\sim\Gamma}[\|X-Y\|^2]$$

A current state of affairs

[Constrained sampling]

Algs

Ball walk, Hit-and-Run

Metrics

TV, χ^2

Tools

Conductance

Fundamental gap here?

[Unconstrained sampling]

Algs

Langevin-based

Metrics

$\mathscr{R}_q$

Tools

Wasserstein calculus, optimal transport, Markov semigroup theory, interpolation method, Girsanov's argument, Shifted composition rule,...........

Let's bridge this gap

[Constrained sampling]

Algs

New sampler

Metrics

$\mathscr{R}_q$ (and $\mathscr{R}_\infty$ in fact)

Tools

Continuous interpolation via a forward/backward SDE

Can borrow these techniques!

[Unconstrained sampling]

Algs

Langevin-based

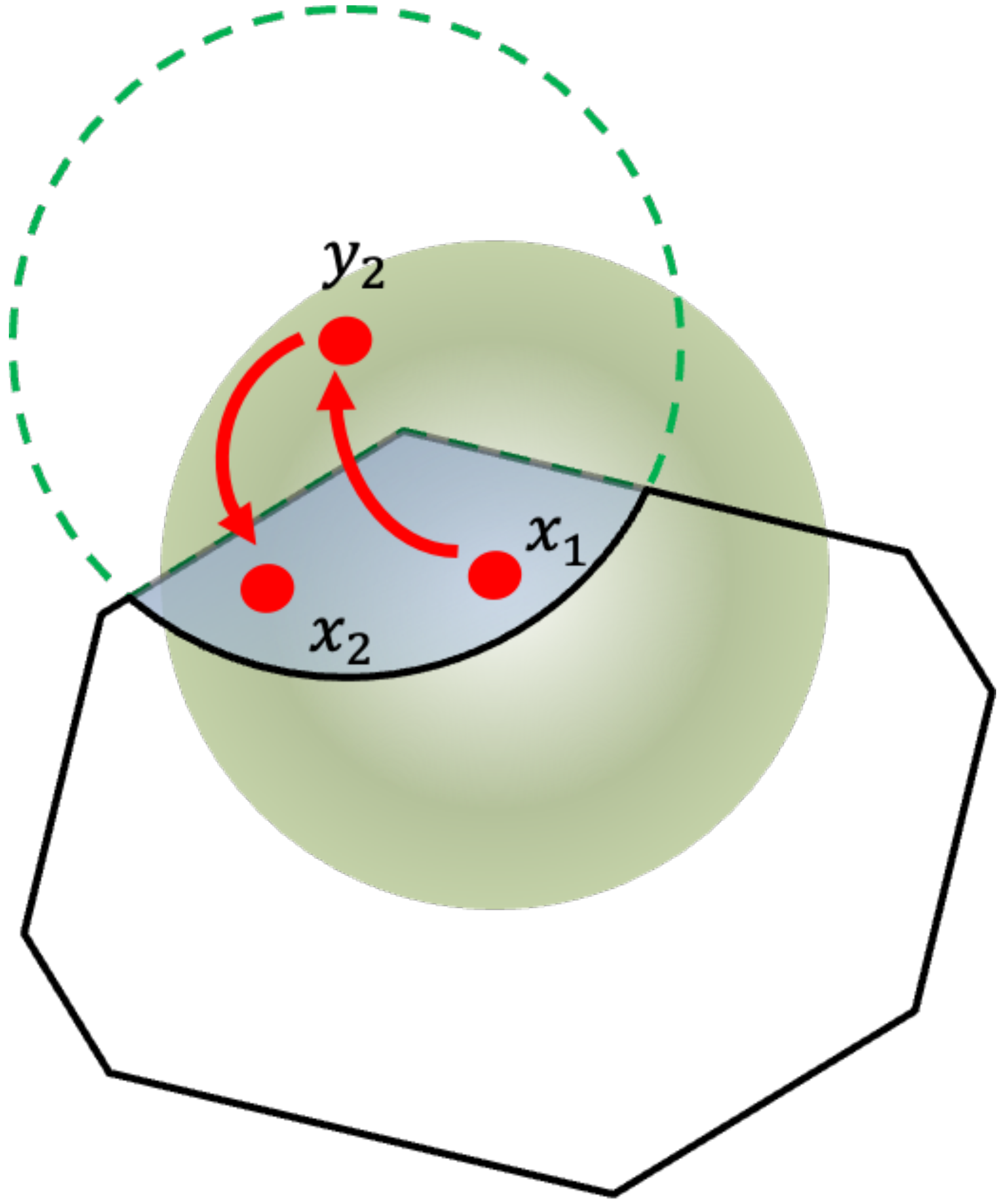
Metrics

$\mathscr{R}_q$

Tools

Wasserstein calculus, optimal transport, Markov semigroup theory, interpolation method, Girsanov's argument, Shifted composition rule,..........

In-and-Out



In-and-Out

[Forward] Sample $y_{i+1} \sim \mathsf{N}(x_i, hI_d)$

[Backward] Sample $x_{i+1} \sim \mathsf{N}(y_{i+1}, hI_d)|_K$

* One iteration = forward + backward step

In-and-Out

Input: initial point $x_0 \sim \pi_0$ & convex body $K \subset \mathbb{R}^d$ & threshold N & step size h
Output: x_T

- For $i = 0,\ldots,T$

  1. Sample $y_{i+1} \sim \mathsf{N}(x_i, hI_d) = x_i + \mathsf{N}(0,hI_d)$

  2. Sample $x_{i+1} \sim \mathsf{N}(y_{i+1}, hI_d)|_K$

     [Implementation]

     - $x_{i+1} \sim \mathsf{N}(y_{i+1}, hI_d)$ until $x_{i+1} \in K$

     - If [# attempts $\geq N$], then declare **Failure**

Where does it come from?

Connection to proximal sampler [Lee, Shen, and Tian'21]

Goal: Sample from $\pi(x) \propto \exp(-V(x))$ over $\mathbb{R}^d$

To this end, augment another variable $y \in \mathbb{R}^d$ to consider

$$\pi(x, y) \propto \exp\left(-V(x) - \frac{1}{2h}\|x - y\|^2\right)$$

Algorithm: Repeat

1. Sample $y_{i+1} \sim \pi^{Y|X=x_i}(y) = \mathsf{N}(x_i, hI_d)$
2. Sample $x_{i+1} \sim \pi^{X|Y=y_{i+1}}(x) \propto \exp\left(-V(x) - \frac{1}{2h}\|x - y\|^2\right)$

Where does it come from?

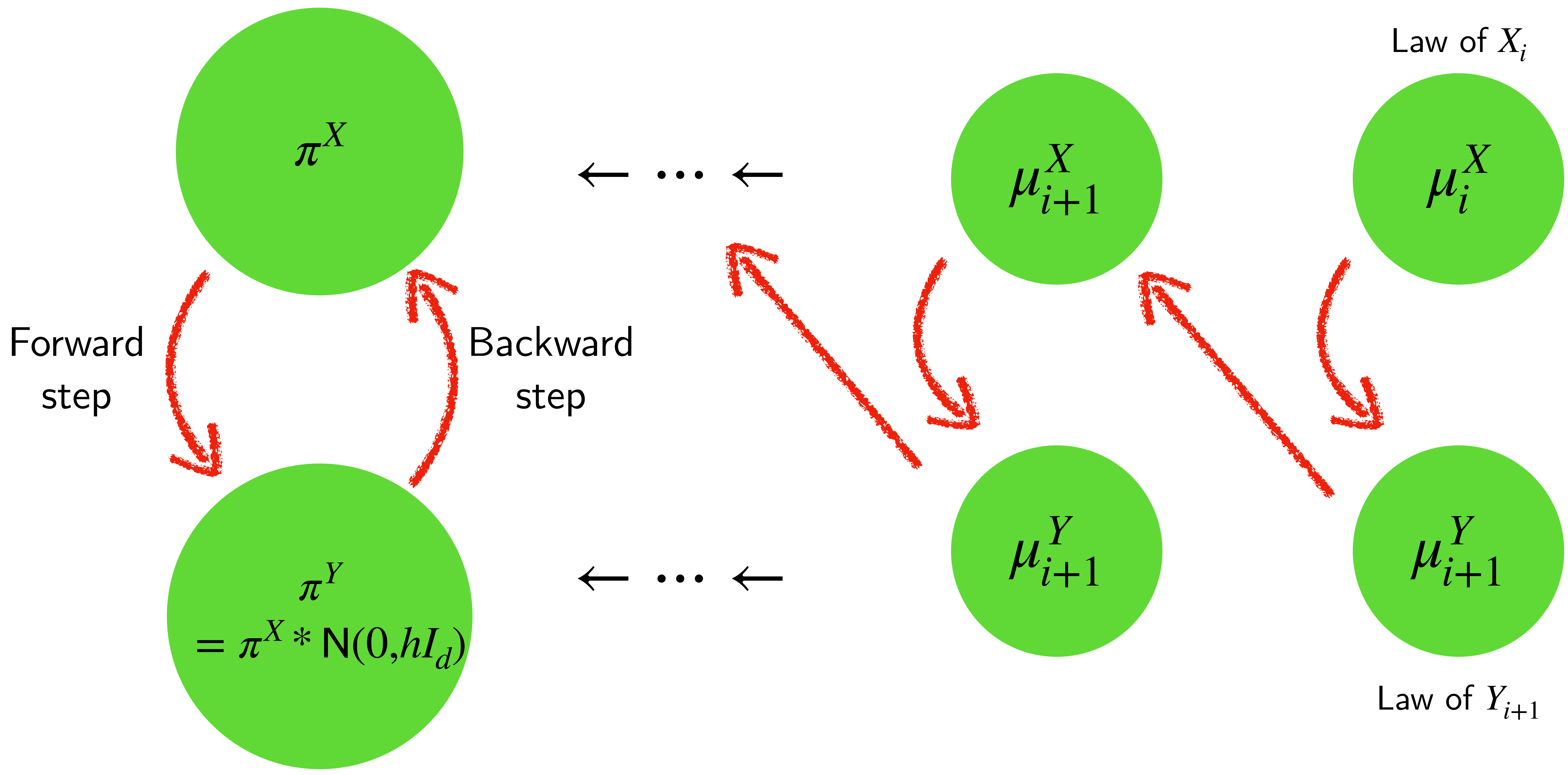
Connection to proximal sampler [Lee, Shen, and Tian'21]

In-and-Out is the Proximal sampler with $\pi(x) \propto 1_K(x)$

Algorithm: Repeat

1. Sample $y_{i+1} \sim \pi^{Y|X=x_i}(y) = \mathsf{N}(x_i, hI_d)$
2. Sample $x_{i+1} \sim \pi^{X|Y=y_{i+1}}(x) \propto \exp\big(-\frac{1}{2h}\|x-y\|^2\big)\,|_K$

Proximal sampler in measure level



Outline of analysis

1. **Contraction** through one-iteration of INO (proximal sampler)

2. **Query complexity** of the implementation for the backward step

Contraction via forward/backward heat-flow

Forward / backward SDE interpretation by [Chen et al.'22]

Forward step: Sample $y_{i+1} \sim \pi^{Y|X=x_i}(y) = \mathsf{N}(x_i, hI_d)$

$$Z_0 \sim \mathrm{law}(X_i) \longrightarrow \mathrm{d}Z_t = \mathrm{d}B_t \text{ for } t \in [0,h] \longrightarrow Z_h \sim \mathrm{law}(Y_{i+1})$$

Contraction via forward/backward heat-flow

Forward / backward SDE interpretation by [Chen et al.'22]

Backward step: Sample $x_{i+1} \sim \pi^{X|Y=y_{i+1}}(x) \propto \exp\left(-V(x) - \frac{1}{2h}\|x - y_{i+1}\|^2\right)$

$Z_0 \sim \mathrm{law}(X_i)$ → $\mathrm{d}Z_t = \mathrm{d}B_t$ for $t \in [0,h]$ → $Z_h \sim \mathrm{law}(Y_{i+1})$

$Z_h^{\leftarrow} \sim \mathrm{law}(X_{i+1})$ ← $\mathrm{d}Z_t^{\leftarrow} = \nabla \log(\pi^X Q_{h-t})(Z_t^{\leftarrow})\,\mathrm{d}t + \mathrm{d}B_t$ for $t \in [0,h]$ ← $Z_0^{\leftarrow} = Z_h$

Contraction via forward/backward heat-flow

Q. Benefits of introducing an SDE representation?

A1. Avoid discretization-analysis
A2. Use tools from Markov semigroup theory

Contraction via forward/backward heat-flow

Forward / backward SDE interpretation by [Chen et al.'22]

Using Markov semigroup theory + Wasserstein calculus, one can show

Forward SDE: $\mathrm{d}Z_t = \mathrm{d}B_t$ with $Z_0 \sim \mu_i^X \quad \Longrightarrow \quad Z_h \sim \mu_{i+1}^Y$

$$\chi^2(\mu_{i+1}^Y \| \pi^Y) \leq \frac{\chi^2(\mu_i^X \| \pi^X)}{1 + h/C_{\mathsf{PI}}(\pi^X)} \text{ under a Poincaré inequality (PI)}$$

$$\mathsf{KL}(\mu_{i+1}^Y \| \pi^Y) \leq \frac{\mathsf{KL}(\mu_i^X \| \pi^X)}{1 + h/C_{\mathsf{LSI}}(\pi^X)} \text{ under a log-Sobolev inequality (LSI)}$$

Contraction via forward/backward heat-flow

Forward / backward SDE interpretation by [Chen et al.'22]

Using Markov semigroup theory + Wasserstein calculus, one can show

Backward SDE: $\mathrm{d}Z_t^{\leftarrow} = \nabla \log(\pi^X Q_{h-t})(Z_t^{\leftarrow})\,\mathrm{d}t + \mathrm{d}B_t$ with $Z_0^{\leftarrow} \sim \mu_{i+1}^Y \implies \quad Z_h^{\leftarrow} \sim \mu_{i+1}^X$

$$\chi^2(\mu_{i+1}^X \| \pi^X) \le \frac{\chi^2(\mu_{i+1}^Y \| \pi^Y)}{1 + h/C_{\mathsf{PI}}(\pi^X)} \text{ under a Poincaré inequality (PI)}$$

$$\mathsf{KL}(\mu_{i+1}^X \| \pi^X) \le \frac{\mathsf{KL}(\mu_{i+1}^Y \| \pi^Y)}{1 + h/C_{\mathsf{LSI}}(\pi^X)} \text{ under a log-Sobolev inequality (LSI)}$$

Contraction via forward/backward heat-flow

Forward / backward SDE interpretation by [Chen et al.'22]

Composing one forward + backward contraction,

$$\chi^2(\mu_{i+1}^X \| \pi^X) \le \frac{\chi^2(\mu_i^X \| \pi^X)}{\left(1 + h/C_{\mathsf{PI}}(\pi^X)\right)^2} \text{ under a Poincaré inequality (PI)}$$

$$\mathsf{KL}(\mu_{i+1}^X \| \pi^X) \le \frac{\mathsf{KL}(\mu_i^X \| \pi^X)}{\left(1 + h/C_{\mathsf{LSI}}(\pi^X)\right)^2} \text{ under a log-Sobolev inequality (LSI)}$$

* Can still use this result, though π_K is not smooth around ∂K!
(Convolve with $\mathsf{N}(0, \varepsilon I_d)$ & Use the lower-semi continuity of f-divergence as $\varepsilon \to 0$)

* Can be extended to a q-Rényi divergence

Contraction via forward/backward heat-flow

Functional inequalities for π

Poincaré inequality (PI)

$$\mathrm{var}_\pi f \leq C_{\mathsf{PI}}(\pi)\,\mathbb{E}_\pi[\|\nabla f\|^2] \text{ for any smooth } f : \mathbb{R}^d \to \mathbb{R}$$

Log-Sobolev inequality (LSI)

$$\mathrm{Ent}_\pi(f^2) \leq 2C_{\mathsf{LSI}}(\pi)\,\mathbb{E}_\pi[\|\nabla f\|^2] \text{ for any smooth } f : \mathbb{R}^d \to \mathbb{R}$$

Caveat. This depends only on a measure π, not on a Markov chain/kernel

Contraction via forward/backward heat-flow

Known results on log-concave distribution $\pi \propto \exp(-V)$ **over** $\mathbb{R}^d$

1. $C_{\mathsf{PI}}(\pi) \le C_{\mathsf{LSI}}(\pi)$ in general
2. $C_{\mathsf{PI}}(\pi)$
   - $\|\mathsf{Cov}(\pi)\|_{\mathsf{op}} \le C_{\mathsf{PI}}(\pi) \le \psi_d \cdot \|\mathsf{Cov}(\pi)\|_{\mathsf{op}}$
   - KLS conjecture: $\psi_d = \Theta(1)$
   - $\psi_d \lesssim \log d$ [Klartag'23]
3. $C_{\mathsf{LSI}}(\pi) = O(D^2)$ for a log-concave π with support of diameter D

Contraction via forward/backward heat-flow

Relating mixing guarantees to functional inequalities

Theorem. For $\varepsilon \in (0,1)$ and $K \subset B_D(0)$, INO with step-size h and M-warm initial distribution achieves $\mathscr{R}_q(\mu_n \,\|\, \pi_K) \le \varepsilon$ after the following # of iterations:

$$n \asymp \min \begin{cases} q d^2 \|\mathsf{Cov}(\pi_K)\|_{\mathsf{op}} \log \frac{M}{\varepsilon} & \text{for } q \ge 2 \\ q d^2 D^2 \log \frac{\log M}{\varepsilon} & \text{for } q \ge 1 \end{cases}$$

(↑) Substitute the known bounds on $C_{\mathsf{PI}}, C_{\mathsf{LSI}}$ and $h \asymp d^{-2}$

Control over the backward step (RGO)

Rejection sampling for the backward step $x_{i+1} \sim \mathsf{N}(y_{i+1}, hI_d)\,|_K$

[Implementation via rejection sampling]

- $x_{i+1} \sim \mathsf{N}(y_{i+1}, hI_d)$ until $x_{i+1} \in K$

But this is bound to fail

Control over the backward step (RGO)

Rejection sampling for the backward step $x_{i+1} \sim \mathsf{N}(y_{i+1}, hI_d)|_K$

Suppose we're already at stationarity $\pi^X = \mathsf{Unif}(K)$

$$\to \pi^Y = \pi^X * \mathsf{N}(0, hI_d) = \frac{\ell(y)}{\mathsf{vol}(K)}$$

where $\ell(y)$ is a Gaussian version of local conductance [Kannan et al.'97] defined by

$$\ell(y) = \frac{\int_K \exp\left(-\frac{1}{2h}\|x-y\|^2\right) \mathrm{d}y}{\int_{\mathbb{R}^d} \exp\left(-\frac{1}{2h}\|x-y\|^2\right) \mathrm{d}y}$$

→ Simply, the success probability of the rejection sampling at y

Control over the backward step (RGO)

Rejection sampling for the backward step $x_{i+1} \sim \mathsf{N}(y_{i+1}, hI_d)|_K$

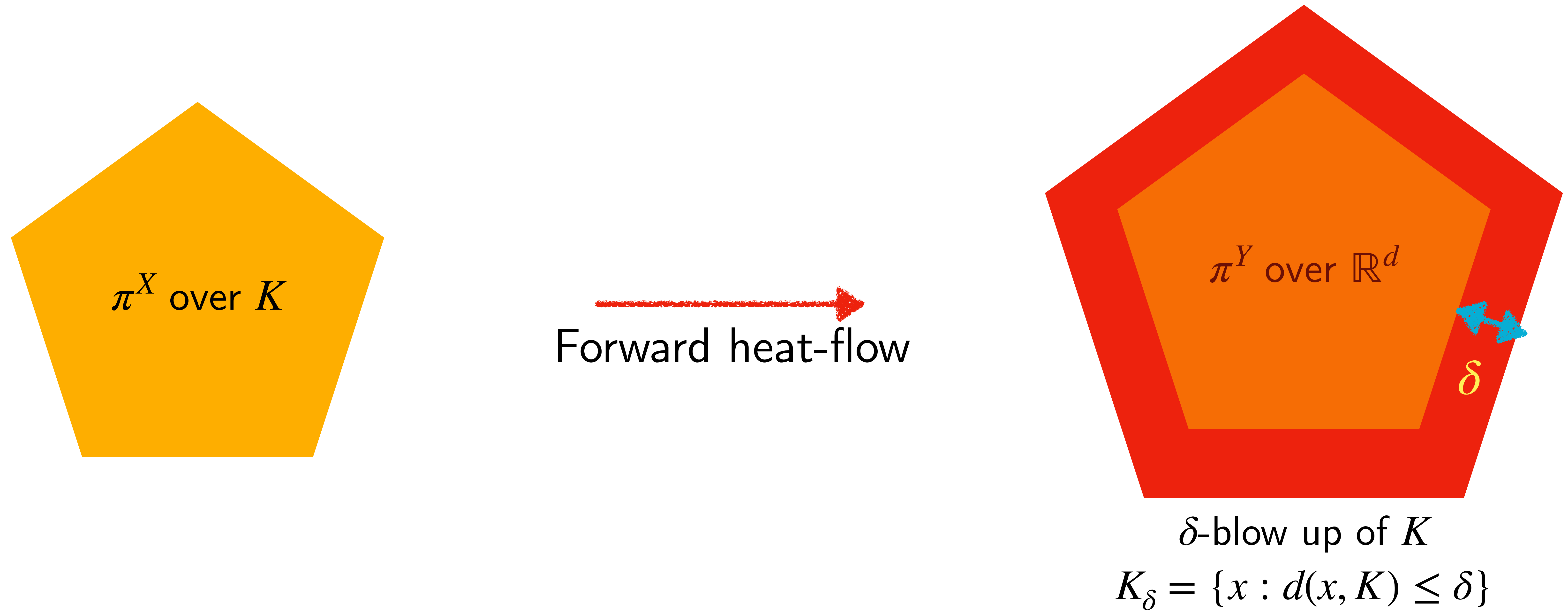
Then the expected number of trials (until success) for one iteration is

$$\mathbb{E}_{\pi^Y}\Big[\frac{1}{\ell(y)}\Big] = \int_{\mathbb{R}^d} \frac{1}{\ell(y)} \frac{\ell(y)}{\mathrm{vol}(K)}\, \mathrm{d}y = \infty$$

Q. Can bypass this issue?

Control over the backward step (RGO)

Rejection sampling for the backward step $x_{i+1} \sim \mathsf{N}(y_{i+1}, hI_d)|_K$

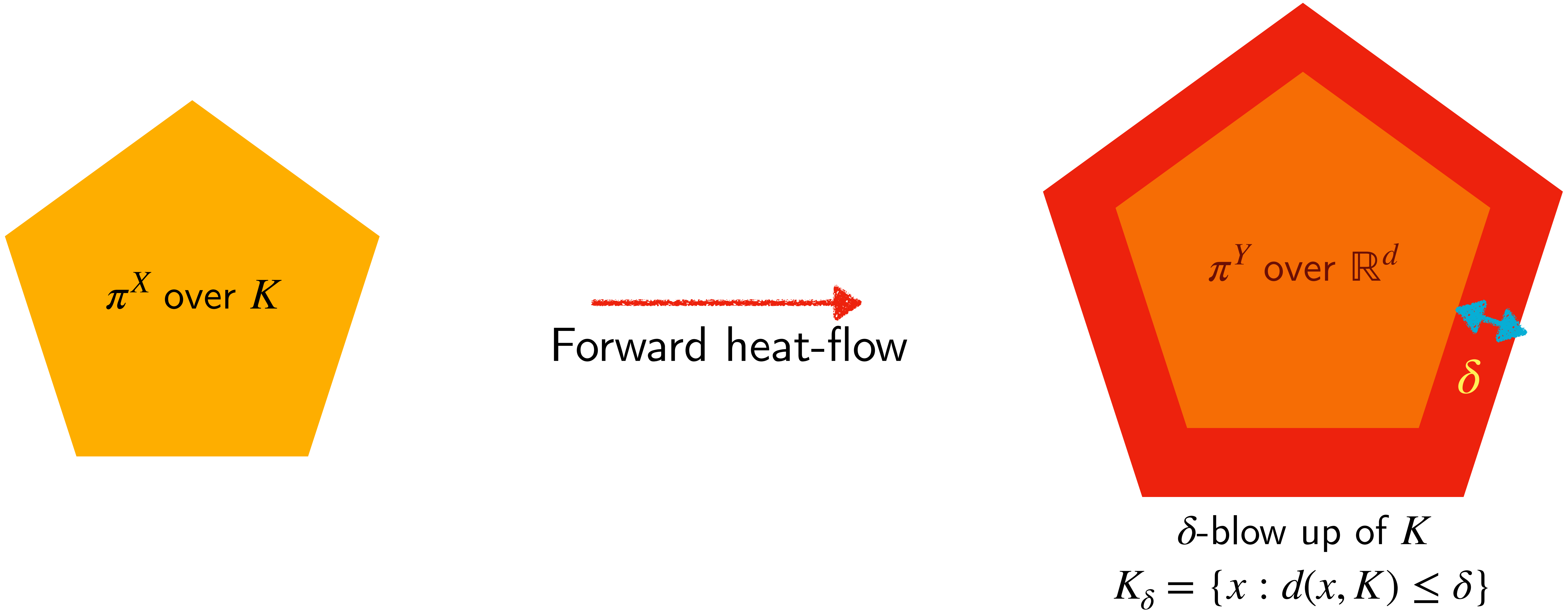


Lemma. $\pi^Y(\mathbb{R}^d \backslash K_\delta) \lesssim \exp(-\Theta(t^2))$ for step-size $h = \Theta(d^{-2})$ and $\delta = t/d$.

$\rightarrow K_\delta$ is sort of "effective domain" of π^Y

Control over the backward step (RGO)

Rejection sampling for the backward step $x_{i+1} \sim \mathrm{N}(y_{i+1}, hI_d)\,|_K$



Insight: Ignore whatever happens outside of this effective domain K_δ

Control over the backward step (RGO)

Rejection sampling for the backward step $x_{i+1} \sim \mathrm{N}(y_{i+1}, hI_d)\,|_K$

Q. **Characteristic** of the complement of the effective domain K_δ?

Proposition. For $y \in K_\delta^c$ with $\delta = t/d$ and $h \asymp d^{-2}$,

$$\ell(y) \le \exp(-\Omega(t^2))\,.$$

Expected #trials from K_δ^c for the rejection sampling $\rightarrow \ell(y)^{-1} \ge \exp(\Omega(t^2))$

∴ Can ignore algorithmic behaviors from K_δ^c by setting a threshold $N = \widetilde{O}\big(\exp(t^2)\big)$
and considering the algorithm as having "failed" if #trials $\ge N$

Control over the backward step (RGO)

Rejection sampling for the backward step $x_{i+1} \sim \mathrm{N}(y_{i+1}, hI_d)|_K$

Theorem. (Complexity of backward step) For failure prob. $\delta \in (0,1)$ and $T \in \mathbb{N}$, there exists suitable choices of parameters h, N such that

1. failure prob. of one backward-step $\leq \delta/T$
2. the expected # of queries per backward-step $\lesssim M \,\mathrm{polylog}(TM/\delta)$

$\therefore$ During T iterations, (1) the total failure prob. is $\leq \delta$, and (2) the total query complexity is $\widetilde{O}\big(MT\,\mathrm{polylog}(1/\delta)\big)$

Guarantee of In-and-Out

Assumption:

1. Access to a membership oracle for a convex body $K \subset \mathbb{R}^d$ with unit ball in K.
2. An initial π_0 is M-warm w.r.t. target $\pi = \mathrm{Unif}(K)$
   - Precisely, $\mathrm{d}\pi_0/\mathrm{d}\pi \le M$ a.s.

Theorem. Given failure prob. $\delta \in (0,1)$, target acc. $\varepsilon \in (0,1)$, and $q \ge 1$, there exists choices of parameters h, N such that with probability $\ge 1 - \delta$,

INO started at π_0 ensures $\mathcal{R}_q(\mathrm{law}(X_n) \| \pi) \le \varepsilon$
after $n = \widetilde{O}\left(qd^2 \|\mathrm{Cov}(\pi)\|_{\mathrm{op}}\, \mathrm{polylog}(M/\delta\varepsilon)\right)$ iterations,
using $\widetilde{O}\left(qMd^2 \|\mathrm{Cov}(\pi)\|_{\mathrm{op}}\, \mathrm{polylog}(1/\delta\varepsilon)\right)$ membership queries in expectation.

Matching results of Ball walk

Theorem. Given failure prob. $\delta \in (0,1)$, target acc. $\varepsilon \in (0,1)$, and $q \geq 1$, there exists choices of parameters h, N such that with probability $\geq 1 - \delta$,

INO started at π_0 ensures $\mathcal{R}_q(\mathrm{law}(X_n) \| \pi) \leq \varepsilon$
by using $\widetilde{O}\big(qMd^2 \|\mathrm{Cov}(\pi)\|_{\mathrm{op}}\, \mathrm{polylog}(1/\delta\varepsilon)\big)$ membership queries in expectation.

Previous best complexity via **Ball walk**:
→ Achieving ε-TV distance from M-warm start needs $O\big(Md^2 \|\mathrm{Cov}(\pi)\|_{\mathrm{op}}\, \mathrm{polylog}(1/\delta\varepsilon)\big)$ queries.

INO recovers the matching result under stronger performance metrics and principled approaches!

Chicken-and-egg problem

Awkward situation...

INO needs $\widetilde{O}\left(qMd^2\|\mathrm{Cov}(\pi)\|_{\mathrm{op}}\,\mathrm{polylog}\frac{1}{\delta\varepsilon}\right)$ queries where $M = \exp\left(\mathscr{R}_\infty(\pi_0\|\pi)\right)$

Observation

Needs $\mathscr{R}_\infty$-warmness to get $\mathscr{R}_q$-result (denote INO: $\mathscr{R}_\infty \to \mathscr{R}_q$).
Same issue with BW (SW + rejection): $\mathscr{R}_\infty \to$ TV.

Q. How to get a warm-start in $\mathscr{R}_\infty$?

Part II - Collaborators

Rényi-infinity constrained sampling with d^3 membership queries

SODA'25

Matthew Zhang
University of Toronto

Warm-start generation

Problem. Let $K \subset \mathbb{R}^d$ be a "well-rounded" convex body (i.e., $\mathbb{E}_\pi[\|X\|^2] = O(d)$) containing a unit ball. Can we generate a warm start X such that

$$\mathcal{R}_\infty(\mathrm{law}(X) \,\|\, \pi) = O(1)?$$

Note) There is a known method for making K well-rounded [Jia et al.'21]

Previous approaches

A common approach is “**annealing**”:

$$\mu_0 \to \mu_1 \to \cdots \to \mu_i \to \mu_{i+1} \to \cdots \to \mu_k \to \pi$$

- μ_0: easy dist. (from which we can easily sample)
- Generate μ_{i+1} starting from μ_i (by some samplers)
- Here, μ_i is a warm start for μ_{i+1} (i.e., μ_i and μ_{i+1} are already close)

Recall $\mathscr{R}_2 = \log(\chi^2 + 1)$

Previous approaches

1. Uniform annealing [Dyer-Frieze-Kannan'89 ~ Kannan-Lovász-Simonovits'97]

$$\mu_i = K \cap (2^{i/d} B_1(0)) \qquad \mathscr{R}_\infty(\mu_i \,\|\, \mu_{i+1}) = O(1)$$

2. Exponential annealing [Lovász and Vempala'06]

$$\mu_i \propto \exp(-a_i^\top x)\,|_K \qquad \mathscr{R}_2(\mu_i \,\|\, \mu_{i+1}) = O(1)$$

3. Gaussian annealing [Cousin and Vempala'18]

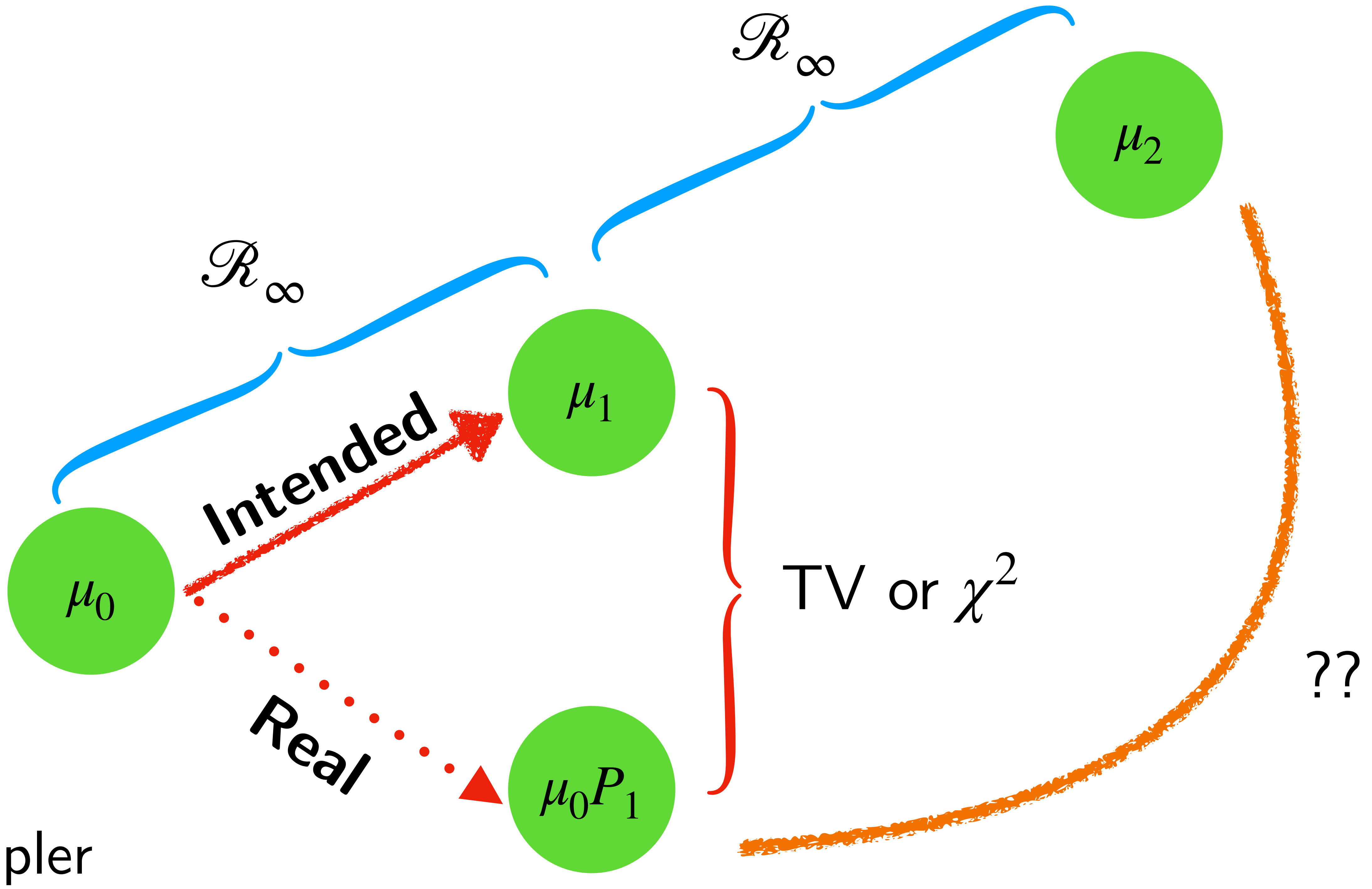
$$\mu_i \propto \exp\Big(-\frac{1}{2\sigma_i^2}\,\|x\|^2\Big)\,|_K \qquad \mathscr{R}_\infty(\mu_i \,\|\, \mu_{i+1}) = O(1)$$

Previous approaches

- Previous works rely on Hit-and-Run or [Speedy walk + rejection sampling] for sampling the annealing distribution μ_i.

However, HAR or SW has guarantees in χ^2 or TV.

Previous approaches



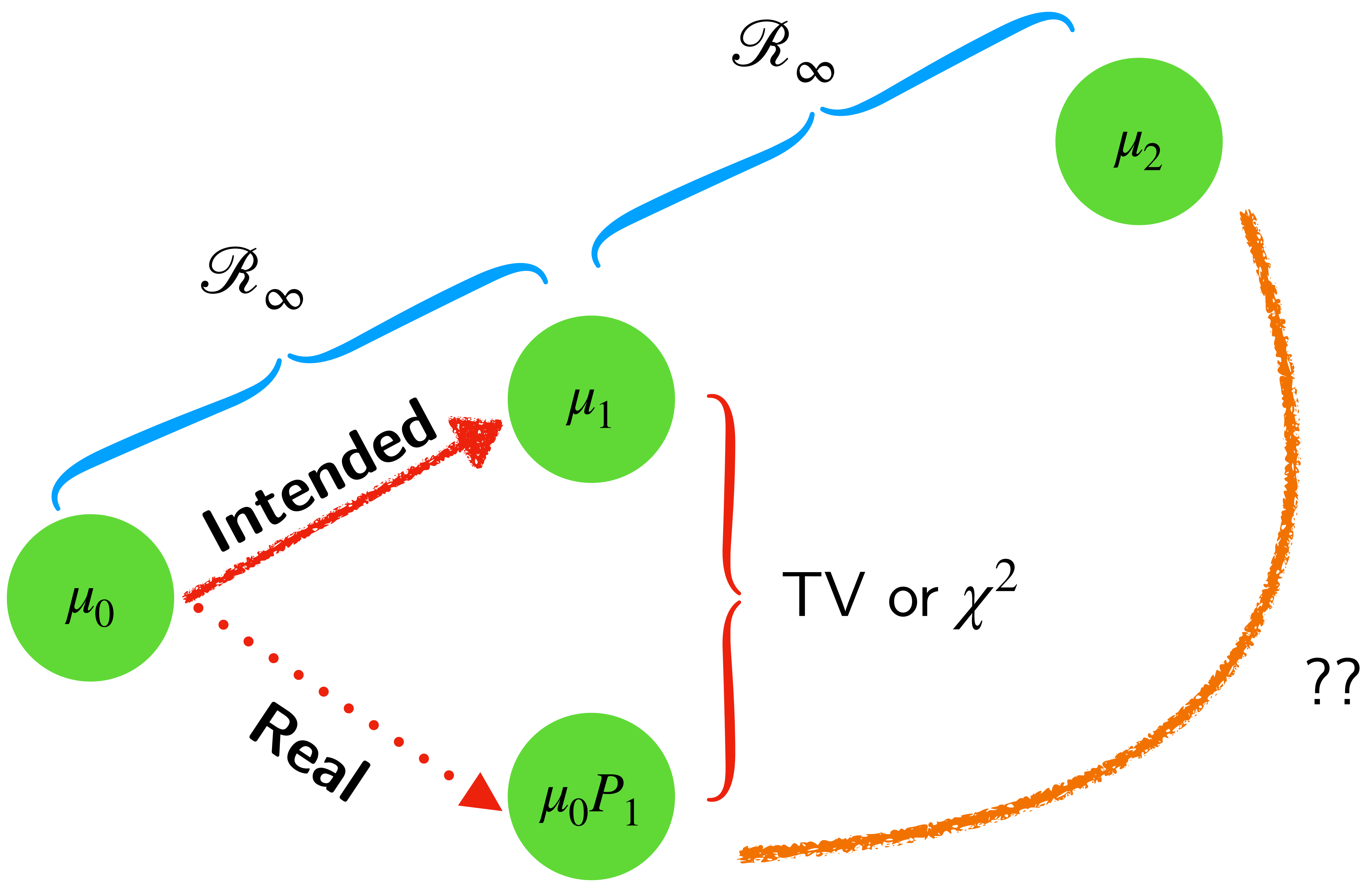
P_1: Markov kernel of the MCMC sampler

MCMC Sampler started at μ_0 will output $X \sim \mu_0 P$ with $\mathrm{TV}(\mu_0 P_1, \mu_1) \leq \varepsilon$

In the next phase, an initial dist. is **in fact** $\mu_0 P_1$, not μ_1.

No triangle inequality coupling TV and $\mathscr{R}_\infty$

Previous approaches



However, the annealing algorithm still proceeds as if the starting distribution is μ_1

Previous works use a coupling argument for analysis, reducing everything to TV.

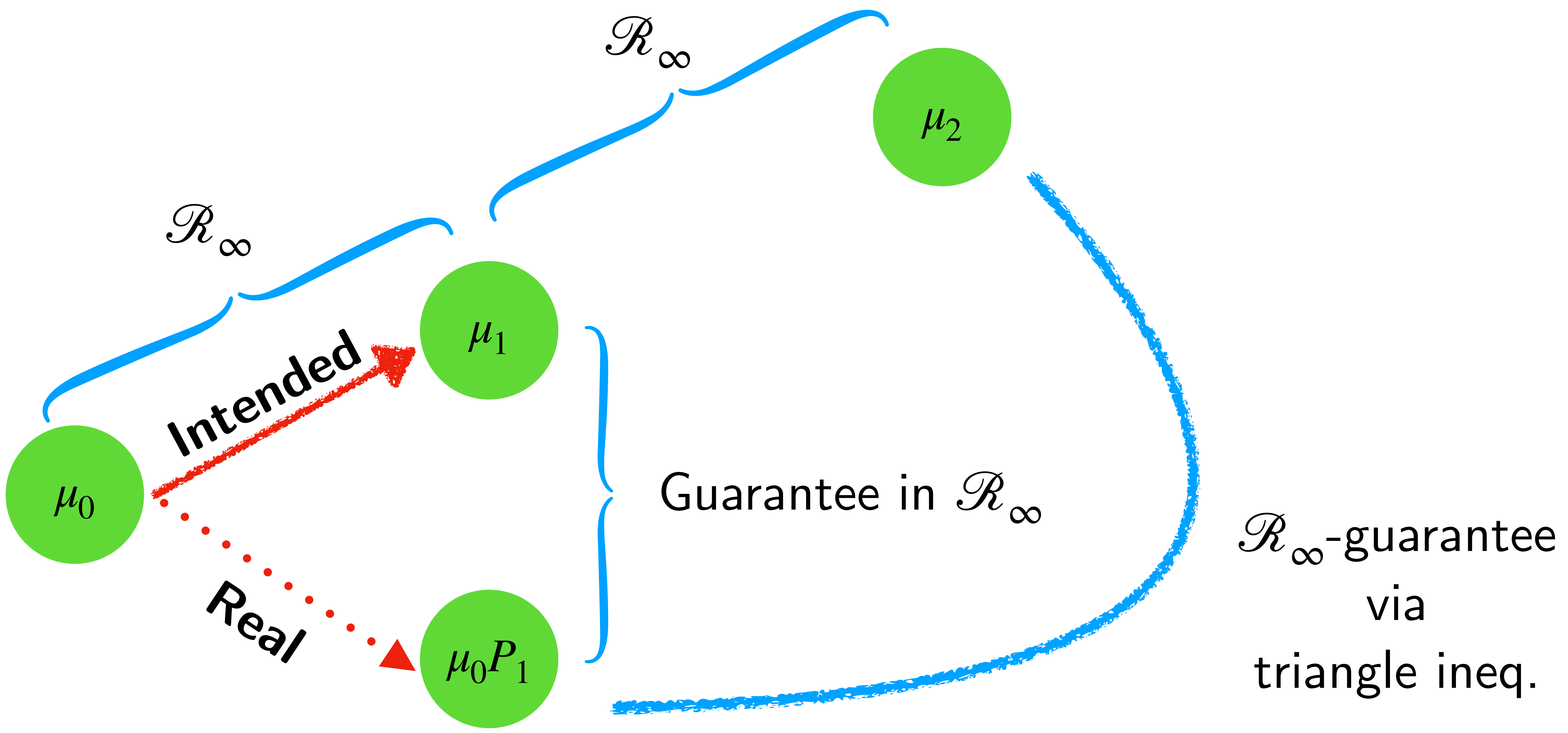
Previous approaches

- Due to inexact error from Markov chains, any guarantee is eventually **collapsed** to TV

    - Previous approaches cannot avoid this "TV-collapse" issue

- If INO uses a warm start generated by this annealing scheme, then its final guarantee ends up collapsing to TV as well

Relay $\mathscr{R}_\infty$-guarantees across annealing?



If a sampler has a $\mathscr{R}_\infty$-guarantee,
then can relay $\mathscr{R}_\infty$-guarantees through the triangle inequality

$\mathscr{R}_\infty$ is difficult

- Prior sampling $\mathscr{R}_q$-guarantees involve a complexity at least linear in q
  - Useless for $\mathscr{R}_\infty$
- A Markov-semigroup approach used for $\mathscr{R}_q$ doesn't go through for $\mathscr{R}_\infty$

In this work, we boost TV $\to \mathscr{R}_\infty$ **without overhead**
via a log-Sobolev inequality (LSI)

Revisit the theory of Markov semigroups

- Let $P : \Omega \times \mathscr{F} \to [0,1]$ be a Markov kernel. Then,

$$\mu P(\,\cdot\,) := \int_\Omega P(\,\cdot \mid x)\, \mu(\mathrm{d}x)$$

$$Pf(x) := \int_\Omega f(y)\, P(\mathrm{d}y \mid x)$$

- Convergence rate is characterized by the **contractivity** of a Markov kernel:

$$\|P\|_{L^p \to L^p} := \sup_{0 \neq f \in L_0^p} \frac{\|Pf\|_{L^p}}{\|f\|_{L^p}}$$

where $L_0^p := \left\{ f : \mathbb{E}_\pi[\,|f|^p\,] < \infty, \mathbb{E}_\pi f = 0 \right\}$.

Revisit the theory of Markov semigroups

The most classical setting is the "$L^2(\pi) \to L^2(\pi)$ contraction"

If $\gamma := \|P\|_{L^2 \to L^2}$, then the so-called **spectral gap** of P is $1 - \gamma$

Q. What about contraction in $L^\infty \to L^\infty$?
(recall $\|f\|_{L^\infty} := \inf\{C : |f| \le C\} = \text{esssup}\, |f|$)

Revisit the theory of Markov semigroups

Q. What about contraction in $L^\infty \to L^\infty$?

Theorem [Rudolf'11]. Let P be a Markov kernel reversible w.r.t. stationary π. Then,

$$\|P^n - 1_\pi\|_{L^\infty \to L^\infty} = 2 \,\mathrm{esssup}_x \,\mathrm{TV}(\delta_x P^n, \pi)$$

where 1_π is the operator defined by $1_\pi(f) := \mathbb{E}_\pi f$.

Convergence from any start implies L^∞-contraction

By substituting $f = \dfrac{\mathrm{d}\mu}{\mathrm{d}\pi} - 1$, one can deduce

$$\mathscr{R}_\infty(\mu P^n \,\|\, \pi) \le \left\| \frac{\mathrm{d}(\mu P^n)}{\mathrm{d}\pi} - 1 \right\|_{L^\infty} \le \left\| \frac{\mathrm{d}\mu}{\mathrm{d}\pi} - 1 \right\|_{L^\infty} \cdot 2\, \mathrm{esssup}_x\, \mathrm{TV}(\delta_x P^n, \pi)$$

$\log(1+x) \le x$

$\therefore$ Uniform TV-bound over any start $x \in \Omega \implies \mathscr{R}_\infty$-bound

Functional inequalities for boosting

Q. When can we bound $\sup_{x\in\Omega} \mathsf{TV}(\delta_x P^n, \pi)$ without huge overhead?

→ (**PI**) ensures an exponential contraction in χ^2 such as

$$\chi^2(\delta_x P^n \,\|\, \pi) \lesssim \exp\Big(-\frac{n}{C_{\mathsf{PI}}(\pi)}\Big)\,\chi^2(\delta_x P^1 \,\|\, \pi)$$

→ (**LSI**) ensures an exponential contraction in KL such as

$$\mathsf{KL}(\delta_x P^n \,\|\, \pi) \lesssim \exp\Big(-\frac{n}{C_{\mathsf{LSI}}(\pi)}\Big)\,\mathsf{KL}(\delta_x P^1 \,\|\, \pi)$$

Functional inequalities for boosting

Q. When can we bound $\sup_{x\in\Omega} \mathrm{TV}(\delta_x P^n, \pi)$ without huge overhead?

Recall $2\,\mathrm{TV}^2 \le \mathrm{KL} \le \log(1+\chi^2) \le \chi^2$. In general,

$$\mathrm{KL}(\delta_x P \,\|\, \pi) = \mathrm{poly}(d)$$
$$\chi^2(\delta_x P \,\|\, \pi) = \exp(\mathrm{poly}(d))$$

Under (**PI**), the convergence rate would have the overhead of $\log \chi_0^2 = \mathrm{poly}(d)$
Under (**LSI**), the convergence rate would have the overhead of $\log \mathrm{KL}_0 = \mathrm{polylog}(d)$

LSI can provide $\mathscr{R}_\infty$-guarantee only with polylog overhead!

Annealing through Gaussians

Work with the **Gaussian cooling** [Cousin and Vempala'18]:

$$\mu_0 \to \cdots \to \mu_i \to \mu_{i+1} \to \cdots \to \mu_k \to \pi \quad \text{with} \quad \mu_i = \mathsf{N}(0, \sigma_i^2 I_d)\,|_K \text{ and } \pi = \mathsf{Unif}(K)$$

→ Need a sampler for a truncated Gaussian

Proximal sampler once again!

Sampling from a truncated Gaussian

Proximal sampler for a truncated Gaussian

$$\pi(x, y) \propto \exp\Big(-V(x) - \frac{1}{2h}\|x - y\|^2\Big) \quad \text{with } V(x) = \frac{1}{2\sigma^2}\|x\|^2 \cdot 1_K(x)$$

Algorithm: Repeat

1. Sample $y_{i+1} \sim \pi^{Y|X=x_i}(y) = \mathsf{N}(x_i, hI_d)$

2. Sample $x_{i+1} \sim \pi^{X|Y=y_{i+1}}(x) \propto \exp\Big(-V(x) - \frac{1}{2h}\|x - y\|^2\Big) = \mathsf{N}\Big(\frac{1}{1 + h\sigma^{-2}}\, y_{i+1}, \frac{h}{1 + h\sigma^{-2}}\, I_d\Big)\Big|_K$

Q. What's (1) the convergence rate and (2) query complexity of the backward step?

Uniform ergodicity of proximal sampler

Theorem. Under suitable choices of parameters, for any $x \in K$,

$$\mathscr{R}_q(\delta_x P^n \,\|\, \pi) \leq \varepsilon \text{ for } \pi = \mathsf{N}(0, \sigma^2 I_d)\,|_K$$

after $n = \widetilde{O}\left(q d^2 C_{\mathsf{LSI}}(\pi) \log \frac{\mathrm{poly}(d, D)}{\varepsilon}\right)$ iterations.

Fact 1 [Bakry-Émery]. $C_{\mathsf{LSI}}\big(\mathsf{N}(\mu, \sigma^2 I_d)\big) \leq \sigma^2$
Fact 2 [Bakry-Gentil-Ledoux]. Convex truncation doesn't increase C_{LSI}

$$\therefore\ C_{\mathsf{LSI}}(\pi) \leq \sigma^2$$

Uniform ergodicity of proximal sampler

Theorem. Under suitable choices of parameters, for any $x \in K$,

$$\mathsf{TV}(\delta_x P^n, \pi) \leq \varepsilon \text{ for } \pi = \mathsf{N}(0, \sigma^2 I_d)\,|_K$$

after $n = \widetilde{O}\Big(d^2\sigma^2 \log \frac{\mathrm{poly}(d, D)}{\varepsilon}\Big)$ iterations.

Boost from TV (from any start) $\to \mathscr{R}_\infty$

$$\mathscr{R}_\infty(\mu P^n \,\|\, \pi) \leq \varepsilon \text{ for } \pi = \mathsf{N}(0, \sigma^2 I_d)\,|_K$$

Query complexity of proximal sampler

Use a rejection sampling to implement the backward-step

Theorem. [Complexity] For a well-rdd convex K, failure prob. $\delta \in (0,1)$, target acc. $\varepsilon \in (0,1)$, $\exists$ parameters h, N such that with probability $\geq 1-\delta$,

the Proximal sampler with M-warm start ensures $\mathcal{R}_\infty(\mathrm{law}(X_n)\|\pi) \leq \varepsilon$ by using $\widetilde{O}\left(Md^2\sigma^2 \operatorname{polylog}\frac{D}{\delta\varepsilon}\right)$ membership queries in expectation.

Annealing through Gaussians

Employ the proximal sampler within the **Gaussian cooling** [Cousin and Vempala'18]:

$$\mu_0 \to \cdots \to \mu_i \to \mu_{i+1} \to \cdots \to \mu_k \to \pi \quad \text{with} \quad \mu_i = \mathsf{N}(0, \sigma_i^2 I_d)\,|_K \text{ and } \pi = \mathsf{Unif}(K)$$

Set $\sigma_0^2 = 1/d$, and update according to

$$\sigma_{i+1}^2 \leftarrow \begin{cases} \sigma_i^2 \left(1 + \frac{1}{d}\right) & \text{if } d^{-1} \le \sigma_i^2 \le 1 \\ \sigma_i^2 \left(1 + \frac{\sigma_i^2}{d}\right) & \text{if } 1 \le \sigma_i^2 \lesssim d \end{cases}$$

Annealing through Gaussians

$$\sigma_{i+1}^2 \leftarrow \begin{cases} \sigma_i^2 \left(1 + \frac{1}{d}\right) & \text{if } d^{-1} \le \sigma_i^2 \le 1 \\ \sigma_i^2 \left(1 + \frac{\sigma_i^2}{d}\right) & \text{if } 1 \le \sigma_i^2 \lesssim d \end{cases}$$

Query complexity of Gaussian sampling from an $O(1)$-warm start: $d^2\sigma^2$

1. During $d^{-1} \le \sigma_i^2 \le 1$, needs $\widetilde{O}(d)$ phases for doubling of σ_i^2 $\rightarrow$ # queries : $d \cdot d^2\sigma^2 \le d^3$
2. During $1 \le \sigma_i^2 \lesssim d$, needs $\widetilde{O}(d/\sigma_i^2)$ phases for doubling $\rightarrow$ # queries : $d/\sigma^2 \cdot d^2\sigma^2 \le d^3$

$\therefore$ Total query complexity through annealing: d^3

Wrap-up in the last phase

Wrap-up for uniform sampling in the last phase

$$\mu_k = \mathsf{N}(0, dI_d)\big|_K \quad \to \quad \pi = \mathsf{Unif}(K)$$

Use the boosting for uniform sampling via LSI

→ INO (or proximal sampler)'s complexity: $\widetilde{O}(d^2 C_{\mathsf{LSI}}) = \widetilde{O}(d^2 D^2)$ in the last phase

Wrap-up in the last phase

Wrap-up for uniform sampling in the last phase

$$\mu_k = \mathsf{N}(0, dI_d)\,|_K \quad \to \quad \pi = \mathsf{Unif}(K)$$

Better way?

Can work with the uniform dist. $\hat{\pi}$ over a **truncated body** $K \cap B_{O(d^{1/2})}(0)$ due to $\mathcal{R}_\infty(\hat{\pi} \,\|\, \pi) = O(1)$

<u>Rationale</u>: A log-concave dist. has a sub-exponential tail

$$P_\pi(\|X - \mu\| \geq t\sqrt{d}) \leq \exp(-t + 1) \text{ when } \mathbb{E}_\pi[\|X\|^2] = O(d)$$

Wrap-up in the last phase

Wrap-up for uniform sampling in the last phase

$$\mu_k = \mathsf{N}(0, dI_d)\,|_K \quad \to \quad \pi = \mathsf{Unif}(K)$$

After truncation by $B_{O(d^{1/2})}(0)$:

$$D = O(d^{1/2}) \text{ so } C_{\mathsf{LSI}}(\hat{\pi}) = O(D^2) = O(d)$$

$\therefore$ INO's complexity is d^3 for uniform sampling in the last phase

Putting together

1. **Annealing through Gaussian**

$$\widetilde{O}\left(d^3 \operatorname{polylog} \frac{D}{\delta}\right) \text{ queries}$$

2. **Uniform sampling in the last phase**

$$\widetilde{O}\left(d^3 \operatorname{polylog} \frac{D}{\delta\varepsilon}\right) \text{ queries}$$

$\therefore \widetilde{O}\left(d^3 \operatorname{polylog} \frac{D}{\delta\varepsilon}\right)$ membership queries for uniform sampling with $\mathscr{R}_\infty$-guarantee

→ Matches the prior best known complexity in TV [Cousin and Vempala'18]